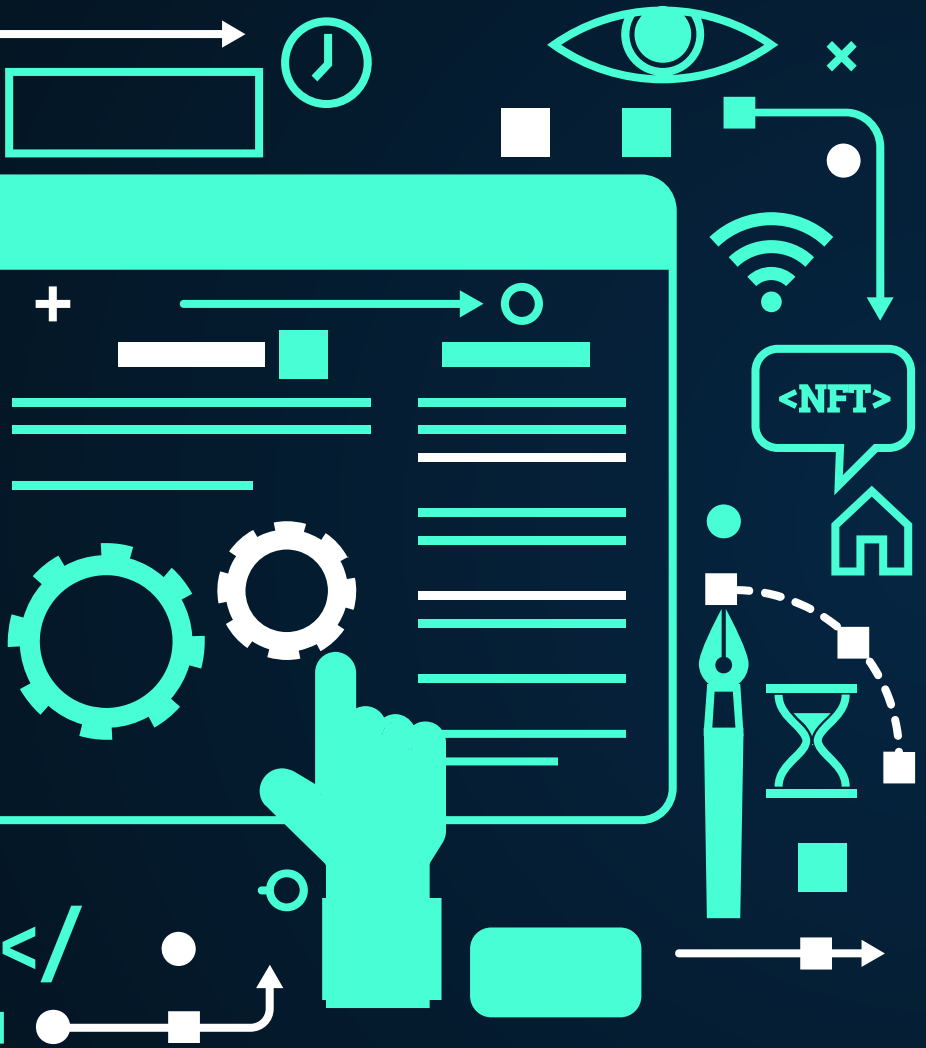


Crypto.io

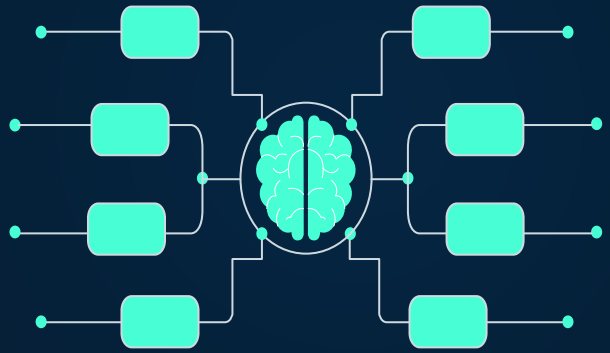
CE/CZ4034 Information Retrieval
(Group 10)

THE TEAM

ISABELLE ONG EE LING (U1921109A)
- Data Exploration
- Classification
- Report
- Slides

JOLENE TAN (U1921255B)
- Crawling
- Classification
- User Interface
- Report
- Slides

TAN YAP SIANG (U1920756H)
- Crawling
- Classification
- User Interface
- Report
- Slides



TAN ZHI WEI SAMUEL (U2020677G)
- Indexing
- User Interface
- Report
- Slides

TEO GUANG XIANG (U2020948F)
- Data Exploration
- Classification
- Report
- Slides

TABLE OF CONTENTS

TABLE OF CONTENTS

Introduction

OUR GOALS

Search Engine

Submit queries to retrieve relevant information

Sentiment Analysis

Provide insights into overall sentiment of NFT community

Time Series

Identify market trends and patterns

NFT Selection

1. Mutant Ape Yacht Club (August 2021)
2. Azuki (January 2022)
3. Bored Ape Yacht Club (April 2021)
4. CloneX (November 2021)
5. Meebits (November 2022)
6. The Potatoz (July 2022)
7. CryptoPunks (June 2017)
8. Phanta Bear (January 2022)
9. MekaVerse (October 2021)
10. Pixelmon (February 2022)

Crawling

SNSCRAPE

Extensive

Allows access to historical tweets

ACCESSIBLE

Does not have rate limits or restrictions

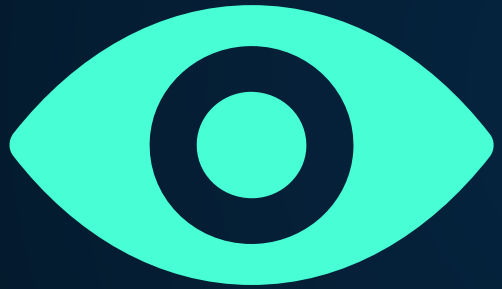
Flexible

Flexible search parameters

Methodology

- ~2000 tweets for each NFT collection
- Release date of the collection
- Equal number of tweets from each quarter
- Eg. Pixelmon (released in February 2022)

Quarter	Number of Tweets
January 2022 to March 2022	449
April 2022 to June 2022	452
July 2022 to September 2022	347
October 2022 to December 2022	467
January 2023 to March 2023	112



Data Exploration

Corpus Information

20295

records

338601

words

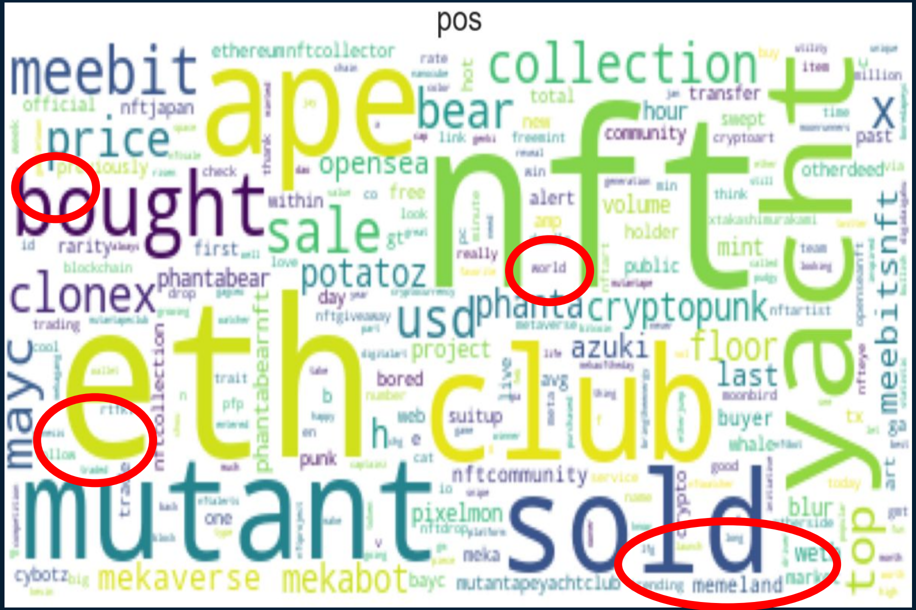
20315

distinct words

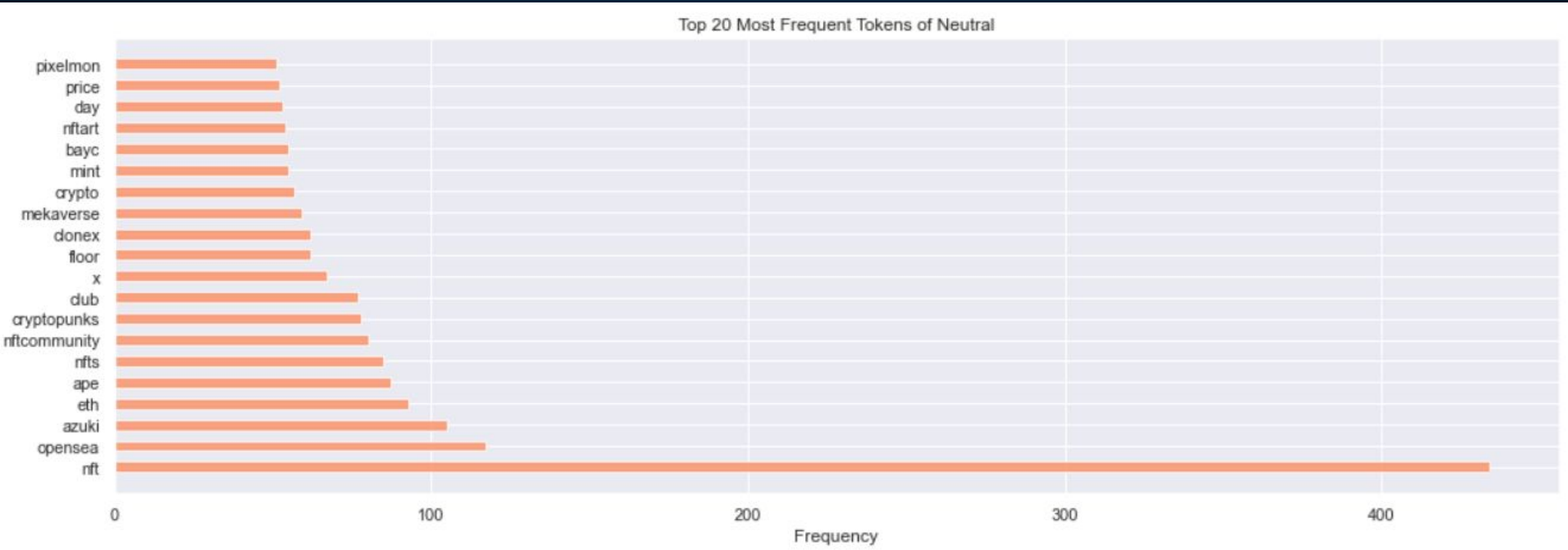
Labeled Dataset

- ~10% (2000 tweets) extracted for manual labeling
- Shuffled
- ~10% (200 tweets) from each marketplace extracted
- Conflicting sentiment labels re-examined

WordCloud


pos
meebit
ape
nft
yacht
collection
bear
eth
club
mutant
sold
bought
price
clonex
sale
opensea
potatoz
usd
cryptopunk
mayc
azuki
floor
last
mekaverse
mekabot
pixelmon
nftcommunity
mutantapeyachtclub
memeland
phantabear
phantabearnft
meebitsnft
top
project
volume
holder
mint
public
buyer
whale
blur
alert
free
official
nftjapan
ethereumnftcollector
transfer
hour
community
otherdeed
total
first
rarity
blockchain
trading
bored
suitup
punk
cybotz
bayc
neka
crypto

Top 20 Tokens



Data Preprocessing

Techniques

- Removal of Non-English and Duplicated Tweets
- Stopword Removal and Tokenization
- Noise Removal
- Case Folding and Lemmatization

Metrics	Before Preprocessing	After Preprocessing
Time	9.47s	1.73s
Accuracy	0.82	0.83
Precision	0.75	0.73
Recall	0.63	0.65
F1-score	0.64	0.67

Data Augmentation

Experiments

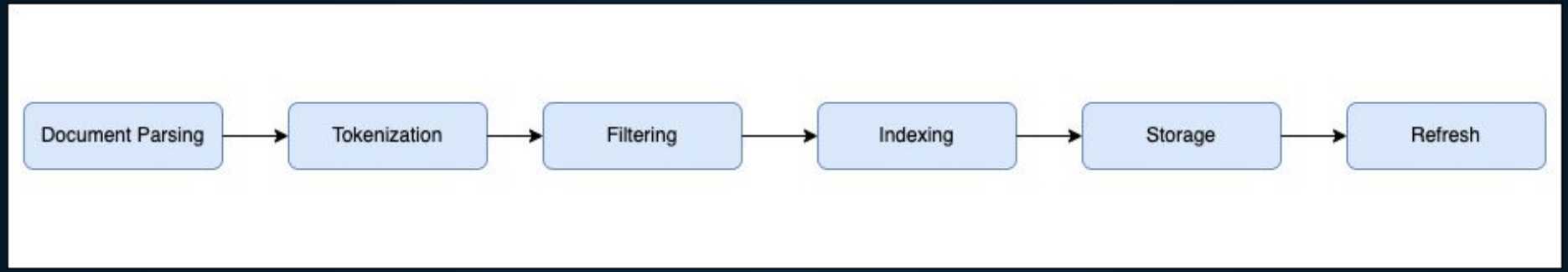
- Synonym Augmentor
- Embedding Augmentor
- Stacking of Synonym + Embedding Augmentors
- Stacking of Embedding + Synonym Augmentors

Evaluation

	Baseline	Synonym	Embedding	Synonym + Embedding	Embedding + Synonym
Count (neg)	88	352	176	704	704
Accuracy	0.83	0.81	0.82	0.82	0.80
F1-score (neg)	0.51	0.47	0.46	0.55	0.50
F1-score (pos)	0.90	0.89	0.89	0.89	0.88
F1-score (neu)	0.72	0.68	0.70	0.68	0.66
F1-score	0.71	0.68	0.68	0.71	0.68

Indexing

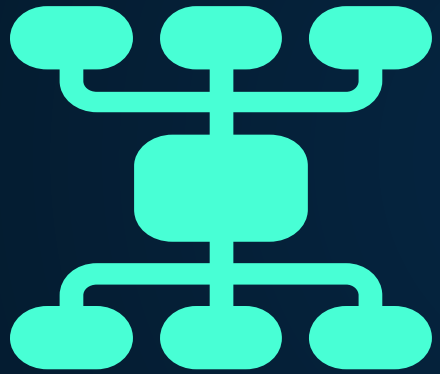
Elasticsearch Indexing



Elasticsearch Querying

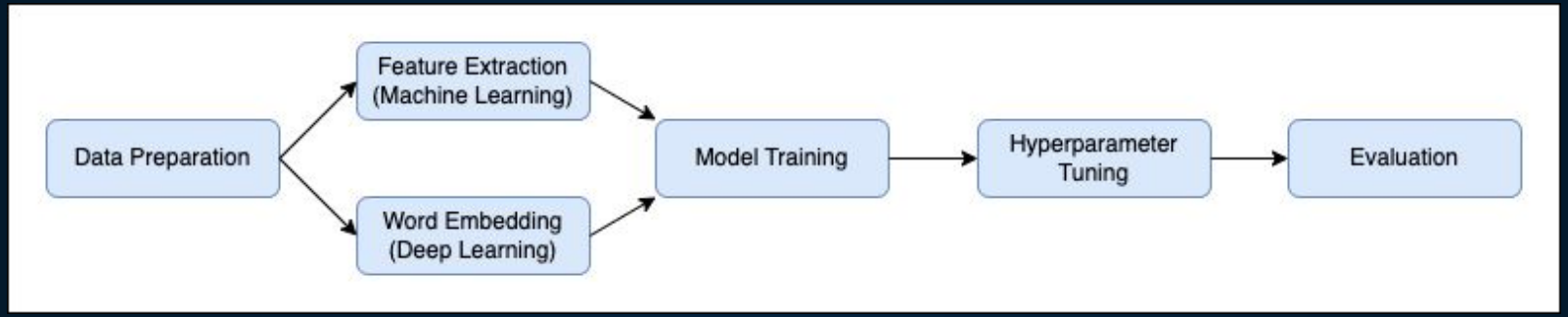
- Elasticsearch Query DSL
- Construct complex queries for searching and filtering data
- Ranked list of documents matching query returned

Query	Time Taken (s)
“azuki best” + NFT: Azuki	0.028
“nft good projects” + NFT: Bored Ape Yacht Club + NFT: CloneX	0.018
“Jaychou” + 3 search results	0.028
“ ” + NFT: Meebits	0.008
“Cryptopunk” + NFT: Meebits	0.013



Classification

Methodology



Ensemble

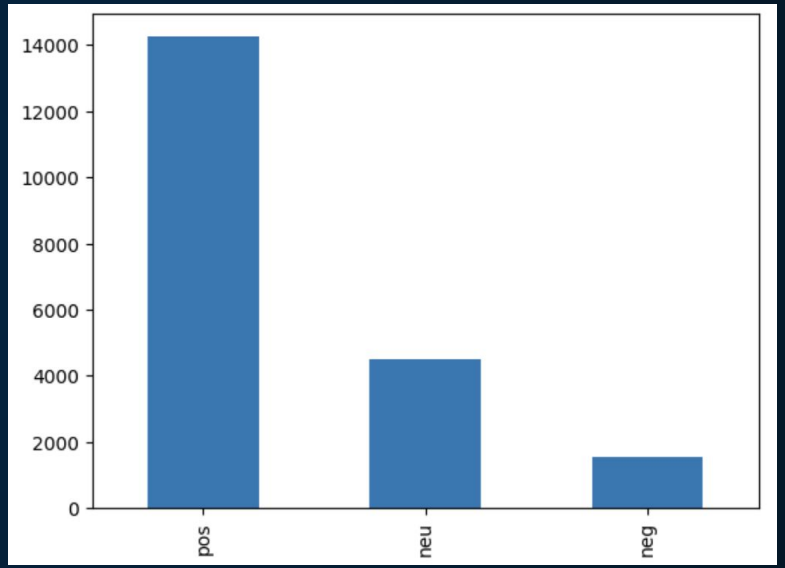
	Training Time	Accuracy	F1-score (neg)	F1-score (pos)	F1-score (neu)	F1-score
XGBoost	2.09s	0.82	0.52	0.90	0.70	0.71
LightGBM	0.76s	0.82	0.56	0.89	0.71	0.72
KNN	0.02s	0.76	0.33	0.85	0.62	0.60
SVM	1.23s	0.82	0.63	0.88	0.65	0.72
Decision Tree	0.05s	0.76	0.58	0.85	0.46	0.63
Ensemble - all	-	0.82	0.59	0.89	0.66	0.71
Ensemble - w/o KNN	-	0.82	0.64	0.89	0.64	0.72

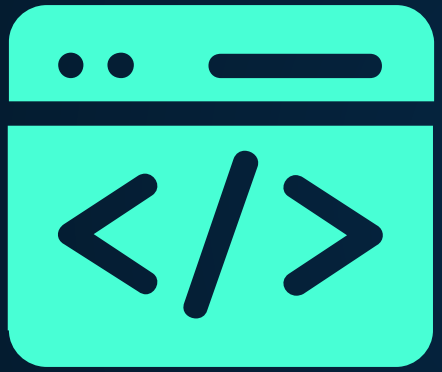
Model Selection

	Training Time	Accuracy	F1-score (neg)	F1-score (pos)	F1-score (neu)	F1-score
Ensemble	-	0.82	0.64	0.89	0.64	0.72
LSTM	38.70s	0.81	0.54	0.88	0.67	0.70
RoBERTa	146.74s	0.86	0.72	0.91	0.75	0.79

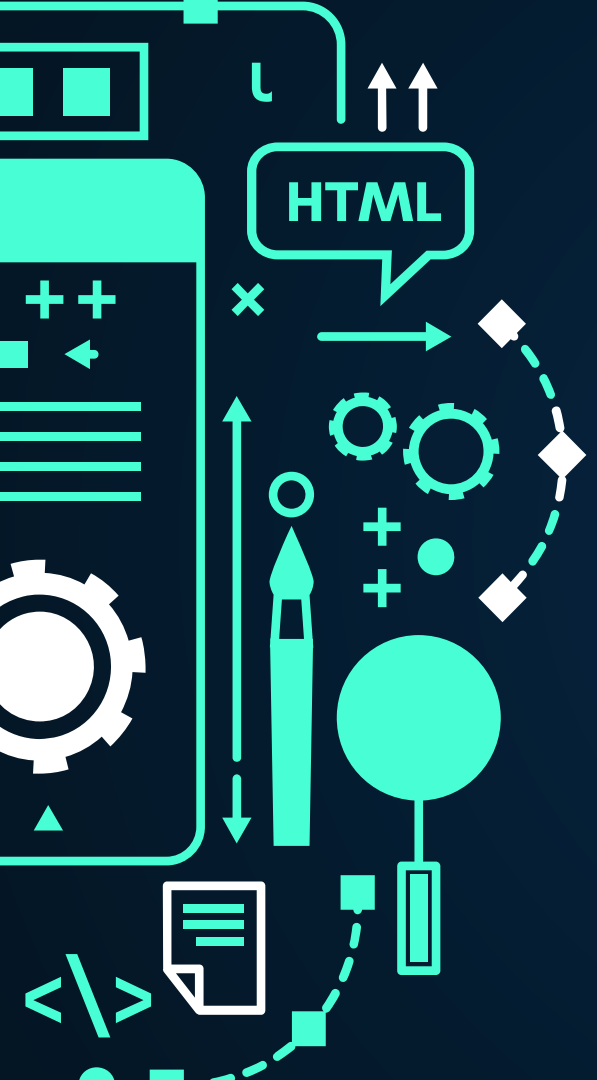
Results

Sentiment	Count
pos	14252
neg	1555
neu	4488





UI Demo



THANK YOU!